CLIPPEDIMAGE= JP404012263A

PAT-NO: JP404012263A

DOCUMENT-IDENTIFIER: JP 04012263 A

TITLE: METHOD AND DEVICE FOR DETERMINING BASE SEQUENCE OF

**NUCLEIC ACID** 

PUBN-DATE: January 16, 1992

INVENTOR-INFORMATION:

**NAME** 

SHIMADA, TAMOTSU

WATABE, KENICHI

NAGAI, KEIICHI

KANBARA, HIDEKI

TAKAHAI, NOBORU

**ASSIGNEE-INFORMATION:** 

**NAME** 

**COUNTRY** 

HITACHI LTD

N/A

HITACHI ELECTRON ENG CO LTD

N/A

APPL-NO: JP02115248 APPL-DATE: May 2, 1990

INT-CL\_(IPC): G01N027/447 US-CL-CURRENT: 324/425

### ABSTRACT:

PURPOSE: To improve the accuracy of the base sequence determination of nucleic acid by correcting the difference in the migration rates between the migration paths of nucleic acid fragments by a computer.

CONSTITUTION: The migration paths 21 are provided in a migration gel 1 sandwiched by glass plates 20 to migrate and separate the nucleic acid fragments. A laser 2 beam is made incident from a side face and the fluorescence from the nucleic acid subjected to fluorescent labeling is detected 3, 4. The data is subjected to correction processing by the computer 23 and is displayed 31. Namely, the time lags of the peaks of the respective bases A, C, G, T determined from count waveforms are corrected by the time lag quantity with predetermined peak intervals by the computer 23. The correction

is infeasible if the correction rate exceeds about 40% of the peak intervals and, there fore, the overlaps and dropouts at the peaks of the respective bases A, C, G, T are measured as the secondary correction and the time lag of the desired base is corrected to minimize these values. The correctable time lag quantity of the secondary correction has the linear relation having the gradient of one base for the first base length and thereafter of one base with each of 60 bases. The accuracy of the base sequence determination is improved by these corrections.

COPYRIGHT: (C)1992,JPO&Japio

## ⑩ 日本国特許庁(JP)

⑩ 特許出願公開

### ⑫ 公 開 特 許 公 報 (A) 平4-12263

®Int. Cl. ⁵ G 01 N 27/447

識別記号

庁内整理番号

個公開 平成 4年(1992) 1月16日

7235-2 J

G 01 N 27/26

3 2 5 325 A 325 E

審査請求 未請求 請求項の数 10 (全 14 頁)

図発明の名称 核酸の塩基配列決定方法及び装置

> @特 願 平2-115248

223出 願 平2(1990)5月2日

⑩発 明 者 H 保

@発 明 者 渡 部 東京都国分寺市東恋ケ窪1丁目280番地 株式会社日立製 作所中央研究所内

健

山口県下松市東豊井794番地 株式会社日立製作所笠戸工 場内

@発 明者 井 啓

東京都国分寺市東恋ケ窪1丁目280番地 株式会社日立製 作所中央研究所内

勿出 願 人 株式会社日立製作所

東京都千代田区神田駿河台4丁目6番地 東京都千代田区大手町2丁目6番2号

の出 願 人 日立電子エンジニアリ ング株式会社

外1名

個代 理 人 最終頁に続く

### 明細書

弁理士 平木 祐輔

1. 発明の名称

核酸の塩基配列決定方法及び装置

- 2. 特許請求の範囲
  - 1. 核酸の断片を泳動分離用ゲルを用いた電気泳 動装置で分子量順に分離、検出する核酸の塩基 配列決定方法において、核酸断片の泳動路間の 泳動速度の差を計算機によって補正することを 特徴とする核酸の塩基配列決定方法。
- 2. 泳動路間の泳動速度の補正が一段階の補正に よるもの又は二段階の補正によるものであるこ とを特徴とする請求項1記載の核酸の塩基配列 决定方法。
- 3. 補正の処理内容を計算機の出力装置に表示す ることを特徴とする請求項Ⅰ記載の核酸の塩基 配列決定方法。
- 4. 塩基配列決定方法の基礎になるピーク間隔を、 計測波形を使って計算機に入力することを特徴 とする請求項1記載の核酸の塩基配列決定方法。
- 5. 塩基配列決定方法の基礎になるピーク間隔を、

計算機によって自動決定することを特徴とする 請求項1記載の核酸の塩基配列決定方法。

- 6. 補正処理するに当りを少なくとも2種類以上 の蛍光体標識した核酸を検出装置で予め処理す ることにより複数の塩基配列を同時に決定する ことを特徴とする核酸の塩基配列決定方法。
- 7. 核酸の断片を泳動分離用ゲルを用いた電気泳 動装置で分子量順に分離、検出する核酸の塩基 配列決定方法において、
- H) 4種類の塩基のうち所望の塩基のピークと、 他の泳動路における他の塩基のそれぞれのビ ークとの間隔を求めるステップ
- ロ ステップので求められたデータに基づき、 時間ずれの量Dを求めるステップ
- ハ ステップ回の時間ずれ量Dに基づき前記所 望の塩基の時間ずれの1次補正を行うステップ なるデータ処理を行う核酸の塩基配列決定方法。
- 8. 核酸の断片を泳動分離用ケルを用いた電気泳 動装置で分子量順に分離、検出する核酸の塩基 配列決定方法において、

- (f) 4種類の泳動路の4種類の塩基A, C, G, T のピークにおいて塩基の重なり及び抜けを求 めるステップ
- (ロ) ステップ(4)で求めた塩基の重なり及び抜け が最小となるように所望の塩基の時間ずれを 補正するための2次補正を行うステップ
- い 2 次補正後の4 種類の塩基のうち所望の塩 基のピークを他の泳動路における他の塩基の それぞれのピークとの間隔を求めるステップ
- (コ) ステップ()で求められたデータに基づき、 時間ずれの量 Dを求めるステップ
- (H) ステップ(二)の時間ずれ量 D に基づき前配所 望の塩基の時間すれの I 次補正を行うステップ

なるデータ処理を行う核酸の塩基配列決定方法。 9. 少なくとも蛍光体を励起するレーザ光源、蛍光検出部および電気泳動分離装置を具備する核酸の塩基配列決定装置において、核酸断片の泳動路間の泳動速度の差を補正するための計算機を設置することを特徴とする核酸の塩基配列決

3

れにより得られる信号は第16図のような波形となり、これらのピークを時系列に読みとって核酸の 塩基配列が決定できる、

この計測波形を 2 次元で表すと、第17図に示すようにアデニン (A)、シトシン (C)、グアニン (G)、チミン (T)ごとにピークが得られる。この場合、T の泳動速度は他の A, C, Gに比べて遅いために、ピークが同一時刻に重なっていて配列決定にミスが生じている。例えば塩基長 212では GとTのピークが重なっている。

このように、泳動路によって核酸断片の泳動速度が異なる原因の一つに泳動ゲルの温度分布の不均一がある。これは装置の使用に当たって、ゲルの作成と装着が手操作であるために、しばしば起こる現象であって、従来の装置の信頼性を低下させる要因になっていた。

## (発明が解決しようとする課題)

上記従来技術では泳動路間の核酸断片の泳動速 度の違いを泳動ゲルを部分的加熱し、その温度分 定装置。

- 10. 少なくとも蛍光体を励起するレーザ光源、蛍光体を標識した遺伝子を検知する蛍光検出部、電気泳動分離装置、核酸断片の泳動路間の泳動速度の差を補正するための計算機を具備してなる遺伝子分析装置。
- 3. 発明の詳細な説明

### (産業上の利用分野)

本発明は、核酸の塩基配列方法及び装置に係り、 特に核酸の塩基配列決定を高精度に行う方法及び 装置に関する。

### 〔従来の技術〕

従来、蛍光標識した核酸断片の検出は、第15図に示す装置で行われていた。この装置ではガラス板20に挟まれた泳動ゲル1に泳動路21を設け、核酸断片を泳動し分離する。

一方、ゲル側面からレーザ2を入射させて全ての泳動路を同時に照射し、蛍光体標識した核酸からの蛍光をイメージ増幅管3と一次元のアレイセンサ4で検出し、計算機22でデータ処理する。こ

4

布を均一化して防止することが行われたが十分な 効果は得られず塩基の配列決定の精度に問題が あった。

本発明は、上記の泳動速度の違いを塩基の配列 決定の原理(同一時刻に4種の塩基のいずれか1 つのピークが存在する。)を使ったソフトウェア によって計算機で補正することにより正確な塩基 の配列決定を行い、本発明を完成するに到った。 〔課題を解決するための手段〕

上記目的を達成するため本発明は、核酸の断片 を泳動分離用ゲルを用いた電気泳動装置で、分子 量順に分離、検出する核酸の塩基配列決定方法に おいて、核酸断片の泳動路間の泳動速度の差を計 算機によって補正することを特徴とするものであ

上記補正は泳動路間の泳動速度の大きさにより 泳動速度の差が小さい場合は1次補正のみの一段 階で行い、泳動速度の差が大きい場合は2次補正 と1次補正を組み合わせた二段階で行う。

1 次補正とは、各塩基 A, C, G, T のピークの

予め決定されているピーク間隔に対する時間ずれ 量を計測し、その時間のずれ量を修正する補正を 意味する。そして、この補正量がピーク間隔の40 %以上を越えると補正が不可能となる。この1次 補正の補正可能な時間ずれ量は第2図に示す通り 最初の補正を区画60塩基長について 0.6塩基、そ の後60塩基毎に 0.6塩基の匂配を有する直線の関 係にある。

次に、2次補正は、上記1次補正において補正不可能な時間ずれ量が生じる場合の適用する補正であって、4種類の泳動路の4種類の塩基 A, C, G, T のピークにおいて塩基の重なり及び抜けを計測し、その重なり及び抜けが最小となるように新望の塩基の時間ずれを修正する補正を意味する。そして、この2次補正の補正可能な時間ずれ量は、2図に示す通り最初の塩基長について1塩基、その後60塩基毎に1塩基の匂配を有する直線の関係にある。

本発明において塩基配列決定方法の基礎となる ピーク間隔は計数波形を使って計算機に入力され、

7 .

21からなる電気泳動分離装置、レーザ光源 2、イメージ増幅管 3 とアレイセンサ 4 からなる蛍光検出部、補正のデーク処理をする計算機23及び補正の情報を出力表示するディスプレイ31を具備するものである。

本発明の泳動路間の塩基の泳動速度の遅いの補 正は次の通り行う。

- (i) 泳動路間の泳動速度の違いを定量化し、その 大きさによって2段階で補正する。
- (2) 計算機によるデータ処理の情況を出力装置に 表示する。
- (3) 2種類以上の核酸断片の混合物を識別する装置の出力データを処理する。
- (4) ピークの時間に隔(ピーク間隔)を正しく設定する。この設定は次の通り行う。
  - ① 計測波形を出力した画面を使って計算機に 入力する。
- ② 検出したビークから計算機を使って自動的 にピーク間隔を計算する。
- (5) 既知の分子量の遺伝子を検出して遺伝子の泳

そのピーク間隔は計算機により自動決定される。 更に本発明において補正の処理内容は計算機の出 力装置に表示される。

本発明においては補正処理するに当り少なくとも2種類以上の蛍光体標識した核酸を検出装置で予め処理することにより2種類以上の蛍光体を標識した複数の核酸の塩基配列を同時に決定することができる。

更に本発明は、少なくとも蛍光体を励起するレーザ光源、蛍光検出部および電気泳動分離器を酸の塩基配列決定装置において、核酸の塩基配列決定差を補正するための計算機を設置することを特徴とする核酸の塩基配子を検徴とする核酸の塩素であり、蛍光体を複識を設置であり、蛍光体を複離装置、核酸断片の泳動速度の差を補正するための計算機を動路間の泳動速度の差を補正する。

前記核酸の塩基配列決定装置は第13図に示す通り、ガラス板20に挟まれた泳動ゲル1及び泳動路

8

動速度の違いを補正する。

更に、計算機による 1 次補正のみの一段階のデータ処理方法の詳細は次の通りである。

即ち、 614 種類の塩基のうち所望の塩基のピークと、他の泳動路における他の塩基のそれぞれのピークとの間隔を求めるステップ、 61ステップ (1) で求められたデータに基づき、時間ずれの量 D を求めるステップ、 61ステップ (1) の時間ずれ量 D に基づき前記所望の塩基の時間ずれの 1 次補正を行うステップ、から成る。

次に、2次補正と1次補正を組み合わせた二段階のデータ処理方法の詳細は次の通りである。

即ち、M4種類の泳動路の4種類の塩基 A、C、G、T のピークにおいて塩基の重なり及び抜けを求めるステップ、ロステップMで求めた塩基の重なり及び抜けが最小となるように所望の塩基の時間ずれを補正するための 2 次補正を行うステップ、ロ2 次補正後の 4 種類の塩基のうち所望の塩基のとークと他の泳動路における他の塩基のそれぞれのピークとの間隔を求めるステップ、ロステップ

いで求められたデータに基づき、時間ずれの量 Dを求めるステップ、(H) ステップ(L) の時間ずれ量 Dに基づき前記所望の塩基の時間ずれの 1 次補正を行うステップ、から成る。

#### 〔実施例1〕

以下、本発明の一実施例を第1図に示したデータ処理フローにより説明する。この処理の入力となるデータは第12図に示したような波形である。この波形データから例えば微分演算によってピークの検出を行った。そして検出したピークの位置を使って、時間ずれの補正を行った。

ところで、DNAの塩基配列決定装置は複数(8種類以上)の試料を同時に分析できるようになっている。一方、本発明の課題になっている時間ずれは、泳動ケルの温度分布の不均一が、その原因と考えられる。

これらのことから、時間ずれは泳動ゲルの中央、 部に供給した試料は小さく、 両端に供給した試料 は大きいという傾向がある。

したがって、時間ずれの小さいデータは1次補

1 1

補正できるようにした。

$$DA = \frac{\sum_{i=1}^{n} p_{i} - \sum_{i=1}^{n} q_{i}}{m + n}$$
 (1)

ここに、

m:ビークA よりも前にあるレーン C, C, T のピークの数。

正だけで処理ができるし、時間ずれの大きいデータは2次補正と1次補正の2段階で補正できるようにした。

ここで時間ずれを定量化したデータの一例を第 2 図に示した。横軸は塩基長で縦軸はピーク間間を塩基長に換算した時の時間ずれの量を示すのの 均力によるものと考えられるので、塩基長のでは 力力によるものと考えられるので、塩基長のでで、塩を 力を減動時間に比例して大きくなる傾向をするで かる。このデータに対する時間ずれの1次ととなるのがである。このデータに対する時間でれの1次とと場合での補正限界線を示すと、例えば塩基長 100の場合、補正可能な時間ずれは1次補正で1塩基、2次補正では3塩基になった。

次に時間ずれの補正方法について具体的に説明する。第2図で説明した様に、時間ずれは塩基長に比例して増加する傾向がある。そこで補正量は60塩基を補正区間として逐次、計算し、その補正量を次の区間に伝播させた(第3図)。これにより、1つの補正区間における時間ずれは小さく、かつ広い範囲の塩基長における大きな時間ずれが

12.

n:ピークA よりも後にあるレーン C, G, T のピークの数。

又、式(I)による結果は、補正する方向も決定でき、正であればピークAIを前に移動し、負であれば後に移動させれば良い。又、この実施例(第4図)ではピークAIの時間ずれについて説明したが、他のピーク C, G, Tも同様の方法で補正した。

この方法では時間ずれの量がピーク間隔の40%を越えると、式(I)から補正する量と方向が正しく計算できないので、これが時間ずれの許容値になる。この許容限界と第3図に示した補正量の伝播方法を考慮すると、時間ずれの1次補正による補正限界が求められる。すなわち、補正可能な時間ずれの量は最初の補正区間60塩基について0.6塩基、その後60塩基毎に0.6塩基の勾配を有する直線になった(第2図)。

次に時間ずれ 2 次補正の具体的な方法を第 5 図で説明する。このデータは D N A の各塩基 A, C, G, T のピークの中で、塩基T のピーク T1, T2 が時間ずれを起こしている場合を示した。この配列

を決定すると、ピークT1とG のピーク位置が重な り(g/t) 、配列の抜け落ち(N) が生じる。ここで 配列決定の原理(ピーク間隔毎に4種類の塩基の いずれか一つが存在する。)を使って、時間ずれ を起こしている塩基の種類を検出した。すなわち、 4種類の塩基をそれぞれ 1 ピーク間隔だけ移動さ せた計16通りの組み合わせについて、ピークの重 なりと抜け落ちが最も少ない組み合わせを求めた。 第 5 図は塩基 A, C, G, ↑ のピークを単独に1つ のピーク間隔だけ移動させた4通りについて示し た。この組み合わせの中で塩基 A, C, Gのピーク を移動しても、ピークの重なりと抜け落ちが生じ ており、↑のピークを移動した場合にピークの重 なりと抜け落ちが解消した。これにより↑のピー クを1つのピーク間隔だけ移動し、さらに時間ず れの1次補正を行うことで、配列が正しく計算で きた。この方法と第3図に示した補正量の伝播方 法を考慮すると、時間ずれの2次補正による補正 限界が求められる。すなわち、補正可能な時間ず れの量は、最初の塩基長で1塩基、以下、塩基長

15

列決定である。そこで、計算機は配列決定中に時間ずれの補正処理の流れに従って第9回に示したように表示するようにした。既に説明したように(第1図)、時間ずれの小さいデータは時間ずれ補正2と補正1の2段階の補正を行う。これを第9回に示したように表示することによって、使用者は装置に供給したDNA試料の時間ずれの情況を知ることができる。

#### (実施例2)

さらに異なる方法(特開平1-261971号)によれば、2種類以上の蛍光体を複識した複数の核酸の塩基配列を、同時に決定できる。この方法で得られる蛍光スペクトルの一例を第10図に示す。スペクトルは4種の塩基 A. C, G. T の時間変化を示しており、左側X1と右側X2は異なる核酸によるものである。またビークに付した番号は核酸に反応させたプライマーからの順番を示した。これによるせたプライマーからの順番を示した。これによれば、核酸X2のビーク72が核酸X1のスペクトルにY72のビークとして表われる。同様に核酸X2のビ

60年に1塩基の勾配を有する直線になった(第2図)。この実施例ではピークTの時間ずれについて説明したが、他のピークの時間ずれ、あるいは複数の塩基の時間ずれについて補正できることは明らかである。

16

ーク75, 76, 77は Y75, Y76, Y77のピークとして 表われる。そこで本発明の配列決定プログラムを 適用するために、次式によるデータ処理が必要に なった。

$$Y_i = X1_i - \alpha X2_i \qquad (2)$$

$$ZZK_i$$

Yı:スペクトルXiの真のデータ

X1::スペクトルX1の計測データ

X2::スペクトルX2の計測データ

α :補正係数

そして、式(2)によって計算した Y. を、本発明による配列決定プログラムで処理することにより、第10図に示した 2 種類の核酸の塩基配列を決定することができた。また、この実施例では 2 種類の核酸の場合について説明したが、さらに多種類の核酸についても、同様の論理で配列決定できることは明らかである。

#### 〔実施例3〕

ところで、本発明による第1次と第2次の時間 ずれ補正を第4図と第5図で説明した。この方法

次にピーク間隔を計算機によって決定する方法 について説明する。第11回に示した計測データに つき、まずピーク検出(例えば微分処理)を行っ た。そして、これらのピーク位置(時間)から、 その差を計算し平均値をピーク間隔とした。

これらピーク間隔は第4図と第5図で説明したように、計測時間に比例して大きくなる。そこで

1 9

ンガープリント)方法に、本発明による配列決定プログラムを適用した。例えばヒトの遺伝子 HLA (Human Leukocyte Antigen) は個人によって部分的に塩基配列が異なる(~10塩基)ことがわかっている。本発明による時間ずれの補正方法は、少なくとも 400塩基以上の配列を決定するので、部分的な配列の違いは時間ずれと判定しない。これにより、異なる個人の間の配列の違い、すなわち個人の識別が可能になった。

### 〔実施例5〕

次に遺伝子診断の装置の実施例を第14図により 説明する。第14図は泳動ゲル1の断面を示したも ので、核酸試料 5 は電気泳動によって断片化して レーザの照射領域を通過する。核酸試料 5 には 試料毎に4 種類の蛍光体を標識して調製したの 式料毎に4 種類の蛍光体を標識して調製したの蛍 光を発する。この蛍光をプリズム 6 で 3 分割した 分割した蛍光は、それぞれフィルタF1、F2、F3を 透過させるようにした。これらの蛍光は極めて 弱なので、イメージ増幅管 3 で増幅し、これをラ 計測時間にともなって増加する塩差長毎にピーク 間隔を設定した。

#### 〔実施例4〕

次に、本発明による配列決定プログラムを遺伝 子の解析に適用した実施例を第12図で説明する。 図の(1)は横軸が遺伝子の大きさを塩基長(キロベ ースペア:KB)を示した。遺伝病(ハンチントン 舞踏病)場合は、2種類の制限酵素H1、H2でDNA を切断すると 4 種のパターン(A, B, C, D)の多型 性が得られる。この検出スペクトル(図の(2)) は 分子量が 2.3KBと8.4KB に共通のピークを持つデ ータが得られる。ここでパターン A~D のピーク に時間ずれがあると、ピークの位置がずれて各ピ 一クの分子量の計算が難しくなる。そこで、本発 明による時間ずれの補正方法を使って、パターン A~D に共通のピークの位置が同じになるように 補正した。すなわち、4種のパターンを平行移動 させる方法で、分子量 2.3KBと8.4KB のピークが 同時に一致する組み合わせを計算した。

さらに核酸の塩基配列を使った個人の識別(フィ

2 0

イン状の一次元アレイセンサ4で電気信号に変換 した。この出力をインターフェース10を介して計 算機11でデータ処理した。

ここで 3 個のアレイセンサの出力を $B_x$ ,  $B_y$ ,  $B_z$  とすると、その色は次式で定量化できる。

$$x = \frac{E_x}{E_x + E_y + E_z}$$

$$y = \frac{E_y}{E_x + E_y + E_z}$$

#### (発明の効果)

本発明において、泳動速度の差を定量化し、2 段階で補正することにより、配列決定の精度が向 上した。更に、基本となるピーク間隔を設定でき るので、配列決定の精度が向上した。

本発明はデータ処理の情況を出力できるので、 使用者は装置の情況を正しく認識できる。

また、本発明の方法は遺伝子のパターンも解析 できるのので遺伝子診断に適用できる。

### 4. 図面の簡単な説明

第1図~第2図は本発明の実施例を説明した図

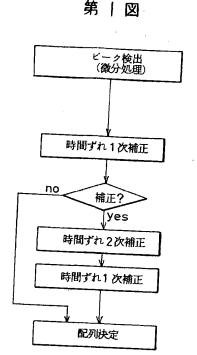
31…ディスプレイ

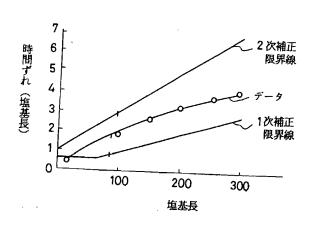
A, C, G, T…DNAの4種類の塩基、X1, X2… 核酸 (2種)、1…電気泳動用ゲル、2…レーザ、3…イメージ増幅管、4…アレイセンサ、5…核酸試料、6…フリズム、10…インターフェイス、21…泳動路、11, 22, 23…計算機、 出願人 株式会社日立製作所 出願人 日立電子エンジニアリング株式会社 代理人 弁理士 平 木 祐 輔 同 弁理士 石 井 貞 次

2 3

2 4

## 第2図



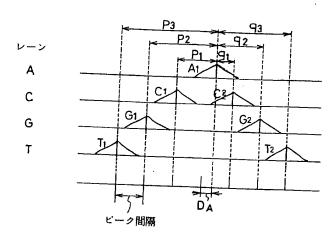


## 第3四

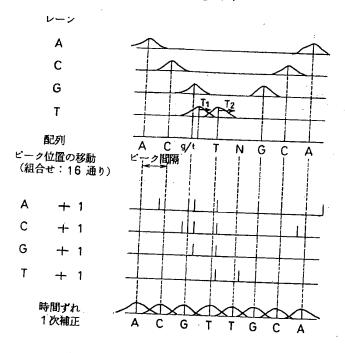


<del>--521--</del>

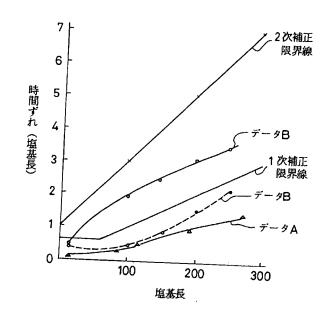
第 4 図



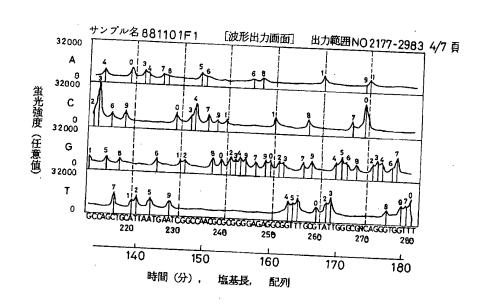
## 第5図



第6図



第7図



## 第8図

## データ解析 1 処理

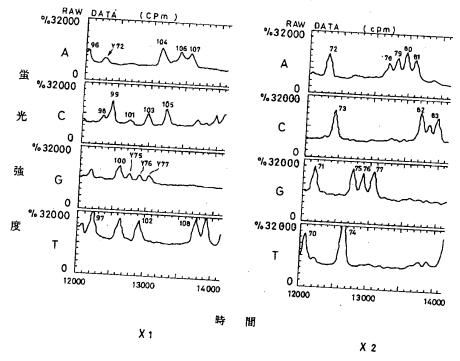
- 1. データ読み込み待ち!!
- 2. ピーク検出待ち!!
- 3. 塩基配列決定待ち!!
- 4. データ書き込み待ち!!

ファイル名: データ確認処理 開始 〈Y/(N)〉 = 入力データ開始ポイント設定(分) = 入力データ終了ポイント設定(分) = データ解析処理 開始 〈Y/(N)〉 =

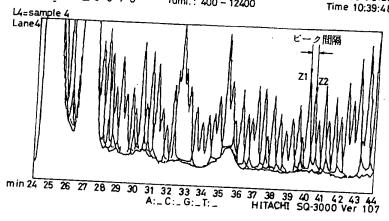
# 第 9 図

シーケンスNO.	データ処理の状態表示
1	時間 <b>ずれ補正 1</b>
2	時間 <b>ずれ補正 2</b>
3	時間ずれ補正 3

# 第 1 〇 図

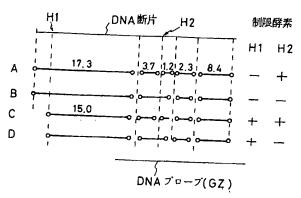


## 第||図

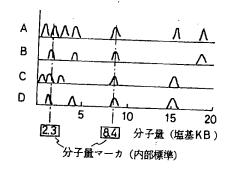


## 第12図

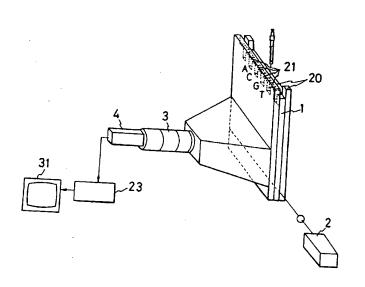
### (1) 遺伝子断片



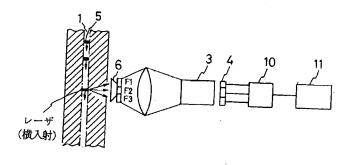
## (2) 検出スペクトル

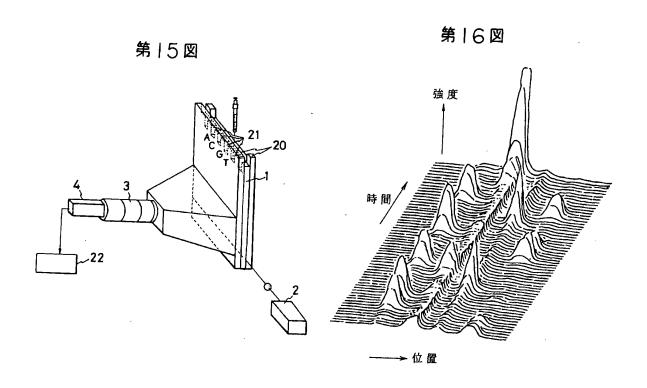


第13図

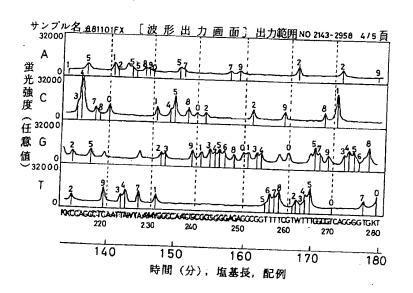


第14図





第17図



第1頁の続き

②発 明 者 神 原 秀 記 東京都国分寺市東恋ケ窪1丁目280番地 株式会社日立製作所基礎研究所内

@発 明 者 髙 杯 昇 東京都千代田区大手町2丁目6番2号 日立電子エンジニアリング株式会社内